DOCUMENT RESUME

ED 097 642 CS 001 383

| | |
|---|---|
| AUTHOR | Williams, Allan R., Jr.; And Others |
| TITLE | Readability of Textual Material--A Survey of the Literature. |
| INSTITUTION | Air Force Human Resources Lab., Lowry AFB, Colo. Technical Training Div. |
| REPORT NO | AFHRL-TR-74-29 |
| PUB DATE | Jul 74 |
| NOTE | 67p.; See related document CS 001 381 |
| EDRS PRICE | MF-$0.75 HC-$3.15 PLUS POSTAGE |
| DESCRIPTORS | Armed Forces; *Literature Reviews; *Readability; Reading; *Reading Comprehension; Reading Instruction; Reading Level; *Reading Materials; Reading Research |

ABSTRACT

This literature review documents the state-of-the-art of readability assessment and indicates directions for future research in readability measurement. The period since 1953 is emphasized, although there is some consideration of earlier work. Primary emphasis within the review is based on readability measurement; however, a final section is included which reviews recent work bearing on the topic of increasing comprehensibility through multimodal presentation methods. Various formulas for calculating readability are presented and placed in historical perspective. The contents include: "Introduction," which presents the scope and organization of the review; "Methods for Measuring Reliability," which discusses rating methods, use tests, readability formulas, early formulas, detailed formulas, recent formulas, cloze procedures, and multimodal presentations; and "Discussion," which discusses the estimation of reading levels, formulas, and future research. An appendix of various readability measures is also included. (WR)

AFHRL-TR-74-29

# READABILITY OF TEXTUAL MATERIAL – A SURVEY OF THE LITERATURE

By

Allan R. Williams, Jr.
Arthur I. Siegel
Applied Psychological Services, Inc.
Wayne, Pennsylvania 19087

James R. Burkett

TECHNICAL TRAINING DIVISION
Lowry Air Force Base, Colorado 80230

July 1974









AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

## NOTICE

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report was submitted by Applied Psychological Services, Inc., Wayne, Pennsylvania 19087, under contract F41609-72-C-0011, work unit 11210403, with the Technical Training Division, Air Force Human Resources Laboratory (AFSC), Lowry Air Force Base, Colorado 80230. Dr. James R. Burkett, Instructional Technology Branch, was the contract monitor.

This report has been reviewed and cleared for open publication and/or public release by the appropriate Office of Information (OI) in accordance with AFR 190-17 and DoDD 5230.9. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS).

This report has been reviewed and is approved.

MARTY R. ROCKWAY, Technical Director
Technical Training Division

Approved for publication.

HAROLD E. FISCHER, Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>AFHRL-TR-74-29 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE *(and Subtitle)*<br>READABILITY OF TEXTUAL MATERIALS -- A SURVEY OF THE LITERATURE | | 5. TYPE OF REPORT & PERIOD COVERED<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Allan R. Williams, Jr.<br>Arthur I. Siegel<br>James R. Burkett | | 8. CONTRACT OR GRANT NUMBER(s)<br>F41609-72-C-0011 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Applied Psychological Services, Inc.<br>Wayne, Pennsylvania 19087 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>62703F<br>11210403 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Hq Air Force Human Resources Laboratory (AFSC)<br>Brooks Air Force Base, Texas | | 12. REPORT DATE<br>July 1974<br>13. NUMBER OF PAGES<br>70 |
| 14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*<br>Technical Training Division<br>Air Force Human Resources Laboratory<br>Lowry Air Force Base, Colorado 80230 | | 15. SECURITY CLASS. *(of this report)*<br>Unclassified<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

readability
reading levels
reading comprehension
training materials
technical writing

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The literature relating to methods of measuring the readability/comprehensibility of textual materials is reviewed and analyzed. Various formulas for calculating readability are presented and placed in historical perspective. The general status of research into the development of readability indices is discussed.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

## SUMMARY

### Problem

It is expected that the mental aptitude and academic achievement levels of enlistees into the United States Air Force may drop as the goal of an all-volunteer military force is approached. Moreover, there is currently a gap between the reading achievement level required to read certain Air Force technical literature and the reading ability levels of enlistees. Accordingly, the Human Resources Laboratory initiated a program to define methods to optimize the matching of technical training materials to the literacy skills of Air Force trainees. This review of the literature relating to methods for determining the readability of textual material represents the first result of this program. The two remaining aspects of the present study are: (1) experimental evaluation of modified training materials, when presented with and without auditory supplementation, and (2) preparation of a training materials modification handbook, which will integrate information from the literature review, the material modification effort, and the experiment.

This literature review is intended to: (1) document the state-of-the-art of readability assessment, and (2) indicate directions for future research in readability measurement.

### Approach

This report selectively reviews the literature relative to readability-comprehensibility measurement. The period since 1953 is emphasized, although there is some consideration of earlier work.

Sources searched for relevant literature included the Psychological Abstracts, from 1950 through 1971, the Technical Abstract Bulletin of the Defense Documentation Center, from 1962 through 1971, and the U. S. Government Research and Development Reports of the Department of Commerce, from 1968 through 1971. The PASAR automated retrieval system of the American Psychological Association was employed to search more completely the literature abstracted in the Psychological Abstracts between 1967 and 1970. Of course, many additional references were found while reading the papers indicated through the above sources.

## Results

Many readability formulas have been derived. The majority are based on linear regression equations relating various observed characteristics of text; i. e., sentence length or word length, to some criterion of comprehension, such as a comprehension test score. All of the formulas are highly intercorrelated and are all undoubtedly highly correlated with cloze score. Cloze score is based on a relatively new procedure in which a judgment of readability is based on the percentage of deleted words in textual material which subjects are able to replace correctly. Cloze score has gained considerable recent acceptance as a readability criterion. But practical considerations may make application of one of the many other available formulas more appropriate in many instances.

## Conclusions

The readability measurement field suffers from the lack of a unifying conceptual or theoretical structure. Clearly, readability is multifactor in nature. The dimensions of readability must be determined with consideration given to variables within both the text and the reader. Upon isolation of these variables, studies to indicate how the variables interact in dynamic reading situations will be required. While the call for a unifying theoretical structure may seem to evade practical issues, it appears that little real progress can be made in readability measurement without such a structure. Specific suggestions for required research are contained within the body of this report.

# TABLE OF CONTENTS

# CHAPTER I

## INTRODUCTION

It is expected that the mental aptitude and academic achievement levels of enlistees into the U. S. Air Force will drop as the goal of an all-volunteer military force is approached (Valentine & Vitola, 1970). Project 100, 000 has also had the effect of lowering the overall academic achievement level of the Air Force. Accordingly, it is expected that efforts to increase the comprehensibility of written materials, both those intended for training purposes and those used on the job, will yield significant advantages. An example of the mismatch between the reading level of military personnel and the readability of the materials they use can be found in the work of Vineberg, Sticht, Taylor, and Caylor (1970), who reported that 75 per cent of a sample of Army Military Occupational Specialties' reading materials were written at a level six to eight school grades higher than the reading level of low-level (Category IV) personnel and four to six grade levels higher than the average reading levels of non-Category IV personnel. The need to match the reading level of Air Force technical literature to the reading level of trainees and job incumbents is quite clear.

### Scope and Organization of this Review

This report reviews the literature relevant to techniques for measuring the readability/comprehensibility of written materials. Use of such techniques could help to improve the intelligibility of Air Force reading materials and by implication reduce the gap between the reading level of the reader and the material he reads. Primary emphasis within the review is based on readability measurement. However, a final section is included which reviews recent work bearing on the topic of increasing comprehensibility through multimodal presentation methods.

The term "readability" may be detined as:

> the sum total (including the interactions) of all those elements within a given piece of printed material that affects the success a group of readers have with it. The success is the extent to which they understand it, read it at an optimum speed, and find it interesting (Dale & Chall, 1949).

Since readability measurement within a training context is the topic of interest here, only the first part of this comprehensive definition--readability as it affects ease of understanding or comprehension--will be considered directly.

Within the measurement concept, "readability formulas," in which attempts are made to predict the readability of written material based on quantitative analysis of the material, will be considered first. These methods have been traditionally based on items such as average word length (in letters), average sentence length (in words), frequency of occurrence of words not appearing on lists of common words, and frequency of occurrence of prepositional phrases and independent clauses. Klare (1963) reports that various reviewers have counted between 29 and 56 such predictive formulas. However, work on this type of formulation has greatly slowed within the past 15 years (Bormuth, 1966; Tannenbaum, & Greenberg, 1968). Studies of readability since that time have tended to concentrate on the "cloze" procedure of W. L. Taylor (1953).

The cloze procedure is the second topic of discussion. In this technique of readability assessment, the percentage of deleted words in a passage correctly filled in by a reader is taken as an index of readability. This procedure overcomes the problem of measuring the readability of technical literature having an unusual, specialized vocabulary which is familiar to individuals within a specific milieu--the precise problem of interest in dealing with technical job-related literature.

# CHAPTER II

## METHODS FOR MEASURING READABILITY

Many approaches are possible to the problem of objectively measuring the readability or comprehensibility of written prose. The most elementary methods; e.g., rating methods and use tests, possess serious drawbacks.

### Rating Methods

In rating methods, judgments of the readability of written manuals are made by samples which are representative of the intended user population or by persons considered to be expert in the covered field. These judgments are necessarily quite subjective, requiring that a large number of raters be employed to "average out" the variability between raters. Raters must also be presented with materials covering a wide range of comprehensibility, so that they may choose the particular materials exhibiting the optimum level of comprehensibility for the intended user population. This procedure may be very expensive to use due to the range of written materials which must be prepared, the necessary size of the rating group, and the effort required to interpret the ratings. Problems may also be encountered in selecting an appropriate rating group and in generalizing from the rating group to the using population.

### Use Tests

The use test method for evaluating readability involves administering comprehension (use) tests to a sample of those for whom the material is intended, after the material has been read by the persons in the sample. High test scores are assumed to be associated with highly readable text, and low scores with less comprehensible text. In addition to the time required to collect and test the sample of users, large amounts of time are required to write tests based on the text.

Standardization of the tests is impossible, a priori, and specific tests may well be much too easy or difficult to rate accurately the text versions of interest. Moreover, if low scores are attained, it is not known whether the low test scores can be attributed to the characteristics of the text itself or, possibly, to the inability of the tested sample to comprehend in general. Finally, there is little guidance available regarding whether the tests should test transfer of factual information, transfer of main points, ability to apply information, or what?

## The Readability Formulas

Quantitative analysis of written text has become the most popular method of assessing readability. The reliability and economy of such methods may be expected to far surpass that of the previously mentioned methods. One reviewer (Klare, 1963) reports that between the publications of the first such formula in 1923 and 1959, over 29 "readability formulas" were developed.

Quantitative analysis of written text has been conducted in hope of finding what determines a readable style. Until very recently, no theoretical model of language behavior was available from which to generate hypotheses. Hence, the method employed in analysis of text has been one of correlation. Characteristically, the text is analyzed and possible factors affecting readability (such as sentence length and vocabulary measures) are conjectured.

A set of readings is then collected and ordered by readability according to a specific criterion (such as reading speed, tested comprehension, judgment). The chosen variables are then measured, their correlations with the readability criteria determined, and regression equations written. Until very recently, linear regression techniques and an additive model have been used exclusively. In generation of the equations, analysts have added additional factors until the increase in predictable variance accounted for by adding another factor was negligible.

In recent years, activity in this area has sharply declined (Tannenberg & Greenbaum, 1968). We may, however, soon witness a renewal of interest in measurement of readability as the science of psycholinguistics grows. Analysis of readability from a theoretical

point of view, e.g., Bormuth (1966), may contribute greatly to the scientific understanding of the written information transfer process. Use of electronic data processing equipment in analyzing readability is also a great aid in overcoming two of the traditional problems in readability analysis: the time and effort required to perform the analysis by hand, and reliability of measurement--both across time and across individuals.

## Early Formulas

The first important attempt to measure objectively readability represented an attempt to aid teachers in choosing appropriate texts for their classes.

In the years around 1920, science teachers at the junior high school level complained that the number of technical and scientific terms present in textbooks intended for classroom use was becoming excessively large. It was becoming necessary to devote large amounts of class time to teaching the meanings of the new vocabulary, lessening the time available to teach scientific concepts. Lively and Pressey (1923) attempted to develop an objective method for determining the vocabulary difficulty, or "burden," of textbooks, so that those books with exceptionally difficult vocabularies could be avoided.

Their measure was based on a simplification of Thorndike's Teacher's Word Book of 10,000 Words, published in 1921. This book lists the frequency of occurrence of the most commonly occurring 10,000 words in the English language. Lively and Pressey used a "weighted median index number" as their measure of vocabulary difficulty. In order to compute the index, a sample of one thousand words evenly distributed through a text was taken. The individual words were found on Thorndike's list, and an index was assigned to each on the basis of its location in the list. For example, those words appearing in the most common thousand (according to Thorndike) received an index number of ten. Those in the second thousand were assigned index number nine, and so on through the ten, thousand-word blocks in the list. Words not appearing on Thorndike's list were assigned an index number of zero.

To compute the vocabulary burden, the median value of the indices of the words sampled from the text was determined, counting each value of zero twice. Substantial agreement was obtained between the judged difficulty of a wide variety of reading selections, ranging from second grade to college level, and the ordering of the selections by the weighted median index number. These findings enabled Lively and Pressey to conclude that the weighted median index number provided an estimate of the vocabulary difficulty of texts. Lively and Pressey were aware of the weaknesses in their readability measure. They pointed out that their index relies on the appropriateness of Thorndike's word count, and this may not be optimal for their intended applications, since Thorndike's count appeared to them to be based largely on materials employing a literary and even poetic vocabulary. They also admitted that larger samples than one thousand words from an entire book may be more appropriate, although they made no effort to evaluate the effects of varying sample sizes.

In addition to its purely historical significance, the readability assessment procedure of Lively and Pressey is significant because it led directly to the first of the "readability formulas" of the modern type--that of Washburne and Vogel (Chall, 1958)--in which various factors correlating with a selected criterion of readability were combined into a single multiple regression equation.

Washburne and Vogel's effort began as one of generating norms for use with the Lively and Pressey method, a need which was pointed out by Lively and Pressey when they presented their initial report. Working in the Winnetka, Illinois, school system, Washburne and Vogel determined the weighted median index numbers of 700 books reported as having been read and liked during the preceding year by at least 25 of the 37,000 students in the school system. Categorization scores comprising a weighted median index number were correlated with median grade levels derived from the paragraph meaning section of the Stanford Achievement Test for those students who reported reading and liking a given book. It was found that a correlation of .80 existed between grade level and number of "zero-value" words present in the tested sample from the corresponding book (Washburne & Vogel, 1926). This information was summarized in the Winnetka Graded Book List, which received considerable use by parents, teachers, and librarians in selecting books to be made available to children in grades 3 through 9 (Chall, 1958).

A need was soon realized for a method of determining the appropriate grade level for books published after the Winnetka list was prepared, without repeating the massive effort employed in developing the original list. Accordingly, Washburne and Vogel selected 150 of the books on the Winnetka list to isolate the internal factors related to reading difficulty; i. e., factors which might be effective in distinguishing those books read by lower grade pupils from those read by students of higher grades. Ten factors were found, and their correlation with grade level was determined. All factors correlated significantly with the criterion of grade level, but only four were subsequently used in deriving a regression equation for measuring readability, since many of the other factors had very high intercorrelations (Vogel & Washburne, 1928).

The readability formula developed by Vogel and Washburne is based on a systematically chosen sample of 1000 words from the book to be tested, which is analyzed as follows:

number of different words appearing ($x_2$)

number of occurrences of prepositions ($x_3$)

number of words not on Thorndike's list of 10,000 ($x_4$)

number of simple sentences in 75 sample sentences ($x_5$)

The difficulty of the book is expressed in terms of grade levels of the Stanford Paragraph Meaning Test ($X_1$). The regression equation derived was:

$$X_1 = .085x_2 + .101x_3 + .604x_4 - .411x_5 + 17.43.$$

This formula correlated .845 with the reading test scores of students who read and liked the respective books during the previous year. It is important to note that the criterion employed here is not strictly comprehensibility of the text. It is confounded by subjective evaluation of the books by students and factors such as subject matter and number of pictures, which are not of direct interest to the question of readability.

In addition to initiating the use of multiple regression equations in the study of readability, Washburne and Vogel's study is significant because of their conceptual approach to the problem. They were the first to: (1) analyze the effects of structural factors in the text, (2) employ an objective criterion of textual difficulty as opposed to qualitative judgment, and (3) describe readability in terms of school grade levels, as opposed to purely relative measures (Chall, 1958).

The readability measures of Lively and Pressey and of Vogel and Washburne are typical of the early studies of readability in that they: (1) measure readability essentially as a function of vocabulary factors, (2) depend heavily on Thorndike's word list, and (3) employ relatively crude criteria of difficulty of text. Other less significant early attempts to measure readability include those of Johnson (1930), Patty and Painter (1931), and Thorndike (1934) himself.

## Detailed Formulas

The readability studies of Ojemann, Dale and Tyler, and Gray and Leary are the most significant of the efforts undertaken between 1934 and 1938, a period characterized by evaluation of larger numbers of factors potentially related to readability, reduced emphasis on word frequency lists, such as that of Thorndike, and concern over more objective readability criteria.

The year 1934 is considered to be the beginning of the trend toward detailed analyses of factors relating to readability, in which a large number of factors, including qualitative ones, were evaluated. The first of these to appear was that of Ojemann (1934), who instituted the use of comprehension test scores as the criterion of readability.

Ojemann collected 16 parent-education passages, each approximately 500 words in length. The passages were read by adults, and comprehension tests were given covering each passage. A reading achievement test was also administered. The grade-level of difficulty of each passage was taken as the mean tested reading level of those people who correctly answered at least half of the comprehension test questions for that selection. After arranging the 16 selections in order of difficulty, Ojemann analyzed their contents for eight sentence factors, six

vocabulary factors, and three qualitative factors. All of the 14 quantitative factors (sequence factors and vocabulary) correlated significantly with the criterion.

Of the sentence factors, three exhibited correlations of .60 or above. These were: number of simple sentences, number of prepositions, and number of prepositions plus infinitives. The five sentence factors having correlations of less than .6 with the criterion of difficulty were: number of complete sentences, number of compound sentences, number of dependent clauses, mean length of dependent clauses, and ratio of total words in independent clauses to the total words in a selection.

All six vocabulary factors correlated .60 or higher with the criterion. These were: percentage of words in Thorndike's first 1000, percentage of words in Thorndike's first 2000, percentage of words known by 70 per cent of sixth-grade pupils, percentage of words known by 90 per cent of sixth-grade pupils, mean difficulty of different words, based on Thorndike's list, and mean difficulty for each word.

No attempt was made to write a regression equation, since the qualitative factors, concreteness versus abstractness of relations, obscurity in expression, and incoherence in expression, appeared to have considerable importance in determining the difficulty of the passages, although they could not be quantified. Instead of a formula, Ojemann presents his 16 tested selections, with their respective values for all the quantitative factors and tests of grade-level difficulty. This type of presentation was intended to allow evaluation of the qualitative factors in testing new passages, in addition to comparing new passages according to the quantitative factors.

In addition to being the first application of comprehension test scores as a criterion of reading difficulty, Ojemann's study was the first to employ adult subjects and adult reading materials. Furthermore, he was the first to demonstrate the importance of qualitative, nonstatistical factors in the determination of readability.

Dale and Tyler (1934) evidenced very great concern over the range of applicability of then-available readability measures. Their interest was in defining those factors determining the difficulty of reading materials dealing with health education for adults of extremely low reading ability. They stated that:

> A critical analysis of the widely varying results of previous studies indicates the impossibility of determining the factors in the reading materials which make them understandable unless the investigations separate the influence of factors within the reading materials from those outside. The reader's interest in the topic treated in the reading matter, his ability to read, the kind of comprehension appropriate to the purposes of the reading matter, and the difficulty of the ideas developed in the reading matter are all factors which greatly affect his comprehension of the material read but are distinct from the characteristics involved in the materials themselves... the various factors not in the reading materials themselves must be controlled in order to determine the effects of factors within the materials (pp. 384-385).

In order to isolate the factors which affect reading difficulty, given a fixed topic, a fixed readership, and a fixed level of comprehension, Dale and Tylor asked adults of low reading ability to read health education articles collected from newspapers, magazines, and books. The readers then completed a multiple-choice comprehension test designed to measure their understanding and retention of the main ideas of the selections.

The type of comprehension test employed was one of the most difficult that could have been used. Questions did not deal with particular items of information which could be remembered with relative ease by the subject. Rather, the subject was required to integrate the entire selection and determine its main point. Additionally, the test questions were written so that, of the five possible responses to each, one option was "best" and the others were characterized by varying degrees of correctness; i.e., accurate statements of secondary points

of the article or slightly inaccurate statements of the points in the article, requiring the subject to know what was covered in the selection, and also what was not covered. Good performance on such a test must have been a formidable task for adults determined to be reading at the third to fifth grade level--the level of these subjects. Not surprisingly, adequate comprehension was not obtained on any of the collected reading materials for determining the correlation between comprehension and the 29 quantitative factors.

To obtain sufficient comprehension, Dale and Tyler found it necessary to write their own, much simplified, test passages. Three principles were found which produced articles with the required ease of reading: (1) use of very basic vocabulary, (2) use of informal style characterized by conversational manner and anecdotal examples, and (3) freedom from digression from the topic of interest.

After rewriting, rescoring, and readministering their materials, Dale and Tyler found that of the 29 factors studied, 10 correlated significantly with comprehension. Of these, three were included in a regression equation for predicting comprehension; i. e., technical words in the passages ($x_2$), the number of nontechnical words not known to 90 per cent of sixth grade pupils ($x_3$) [from an unpublished study by Dale], and the number of indeterminate clauses ($x_4$).

The equation predicted the percentage of adults of reading levels of grade 3 to grade 5 who could understand the main point of a passage ($X_1$) based on the above measures, and took the form

$$X_1 = -9.4x_2 - 0.4x_3 + 2.2x_4 + 114.4.$$

Predicted comprehension correlated .51 with the criterion.

A comprehensive study of factors influencing readability was performed by Gray and Leary (1935). They listed 82 factors relating to readability as drawn from a study of: (1) adults' and children's reading materials, and (2) recommendations of adult students, teachers, and professors of English. Of the 82 factors, 18, such as "image bearing words," "physical and psychic association," etc., were discarded because they seemed to "defy objective measurement." Twenty more

factors were discarded because of their infrequency of occurrence in the sampled materials. Twenty of the remaining 44 factors correlated significantly ($r > .27$) with comprehension test scores of a sample of 756 adult subjects who were given general adult reading materials, both fiction and nonfiction. The subjects were selected in such a way as to be representative of the then current population of adult readers.

When Gray and Leary separated out their upper quartile ("good" readers) and their lower quartile ("poor" readers), the factors correlated differently (by group) with the comprehension test scores. It was found that vocabulary measures were most highly correlated with comprehension scores for poor readers (those achieving comprehension scores in lowest quartile), while readers scoring in highest quartile showed comprehension scores to be most closely related to sentence structure and length.

Gray and Leary provided an entire family of regression equations relating quantifiable factors to predicted comprehension test scores. For low ability readers, as defined by the comprehension test administered, their initial equation included eight factors, and correlated .64 with their criterion. Further analysis showed that considerable reduction in the number of factors included in the equation resulted in minimal change in the multiple correlation with the criterion measure, or in the probable error of the estimate. Nine formulas were presented employing various sets of four of the original eight factors. Multiple correlations computed for these formulas varied from .6350 to .6402, and the probable errors of the formulas varied from .2956 to .2975. These findings allowed Gray and Leary to suggest that in employing their measure of readability, one of the four-factor formulae be employed and that the sole criterion for choosing between them be the ease of measuring the factors required in each formula.

However, the most popular application of Gray and Leary's results has been based on their five-factor formula, which demonstrates a correlation of .64 with the criterion, and the lowest probable error of all (by .0004). This formula takes the form:

$$X_1 = .01029x_2 + .009012x_5 - .02094x_6 - .03313x_7$$
$$-.01485x_8 + 3.774$$

where:

$X_1$ = predicted average comprehension score, along a scale of +4 to -4

$x_2$ = average number of hard words (words not appearing on Dale list of 769 easy words) appearing in samples*

$x_5$ = average number of 1st, 2nd, and 3rd person pronouns appearing in sample

$x_6$ = average sentence length in words

$x_7$ = average percentage of different words in sample

$x_8$ = average number of prepositional phrases appearing in sample

Such an equation obviously lacks broad utility since it is only applicable to low ability readers--the sample on which it is based.

Three hundred fifty books were tested using the five factor formula. The distribution of predicted difficulty of the books was very nearly normal. Five categories of difficulty were defined and labeled "A" ("very easy") through "E" ("very difficult"). Predicted comprehension scores in the "very easy" category ranged from 1.46 to 1.15, and the "very difficult" scores included values from 0.22 to -0.09. Correlation of individuals' comprehension scores with their respective tested reading grade levels indicated that category A materials were written at approximately the 2nd to 3rd grade levels, category B was at roughly the 4th grade level, and category C at the 6th grade to Junior High School level. The two highest difficulty categories did not correlate reliably with reading grade level.

*a 100 word passage from each chapter analyzed.

Other narrower studies of this period are included in the summary table in Appendix A and are briefly discussed by Klare (1963).

## Less Cumbersome Formulas

The next trend in readability research, that toward more efficient and easily applied formulas, saw the development of the two most popular and widely applied readability measures--the Flesch and Dale-Chall formulas.

The appropriateness of searching for efficient formulas was demonstrated by Lorge (1939). He recomputed correlations and regression equations based on the structural variables contained in the five factor formula of Gray and Leary. Lorge used as criterion materials the reading passages of the Standard Test Lessons in Reading of McCall and Crabbs. This is a set of 376 reading passages normed in terms of number of comprehension test questions correctly answered, and related to grade levels of the Thorndike-McCall Reading Scale. Since its initial application by Lorge, this has become the most frequently used and highly respected criterion in studies of reliability (Klare, 1963).

Using the variables of Gray and Leary, Lorge obtained higher correlations with his criterion than had Gray and Leary with any of their formulas. He produced two regression equations, each based on only two of the Gray-Leary variables, $x_2$ and $x_6$, and $x_2$ and $x_8$ (in their notation). Multiple correlations of .7406 and .7456, respectively, were reported. This is the first instance in which over one-half of the variance in a pure measure of comprehension has been accounted for by a readability formula. Lorge attributed his high multiple correlations completely to his use of more adequately standardized criterion materials and his use of a larger sample of criterion materials--376 passages, as opposed to 48 for Gray and Leary.

After study of additional variables, all of which were measures of aspects of vocabulary, Lorge concluded that other variables add insignificantly to the predictive accuracy attainable from vocabulary alone. He suggested that this finding may be attributable to insufficient reliability of available criteria, including that of McCall and Crabbs. But, he contended that improved criterion measures would not negate his thesis that vocabulary is the most important determinant of reading comprehension.

Lorge did not present his own readability formula until 1944 (Lorge, 1944). His formulation included two of the Gray-Leary factors, $x_6$ (average sentence length in words) and $x_8$ (number of prepositional phrases per hundred words), as well as $x_9$ (ratio of hard words to total words in the sample). Lorge categorized these as a sentence structure factor, idea density factor, and vocabulary factor, respectively, and considers them the primary structural elements relative to readability. A "hard word," to Lorge, was one that does not appear on the Dale list of 769 easy words. This list includes words common to Thorndike's first thousand most frequent words and the first thousand most frequent words known to children entering the first grade, from the International Kindergarten List (Gray & Leary, 1935).

Reanalysis of Lorge's data by Dale uncovered certain computational errors, and Lorge's formula was subsequently recomputed in 1948 (Lorge, 1948). The final formula is:

$$C_{50} = .10x_2 + .06x_6 + .10x_8 + 1.99$$

where $C_{50}$ is the reading grade level of individuals answering one-half of the McCall-Crabbs questions on that passage correctly. The obtained predictor-criterion correlation was .67, and addition of other variables raised the correlation to .705.

Lorge is the first to emphasize that readability formulas in general are not to be construed as prescriptions for writing, but are only usable as an approximate of the reading difficulty of a passage.

Rudolf Flesch was strongly dissatisfied with the then available measures of readability as they could be applied to adult reading materials (Flesch, 1943a). Readability formulas, when applied to adult literature, such as magazines, tended to rank the literature in orders far different from the ordering that was produced by judgment or knowledge of the educational levels of readers of each magazine. Flesch thought that this misranking was due to excessive emphasis on range of vocabulary and insufficient stress on other factors.

Flesch hypothesized three types of factors that should highly influence readability for adults. First, he suggested that sentence length should be an important correlate of adult reading difficulty, although it did not appear to be for children. He found support for this idea in the work of Gray and Leary (1935), who found sentence length to be an important variable associated with readability for good readers only. Second, Flesch hypothesized that abstractness was correlated with readability for adults. Third, he thought that the reader's interest in the topic of a reading passage should influence its readability.

Based on Lorge's data employing the McCall-Crabbs Standard Test Lessons in Reading, Flesch developed his initial formula, in which adult reading material would be assigned to a wide range of school grade levels on the basis of: (1) average sentence length, (2) number of affixed morphemes (affixes and suffixes, with certain exceptions), an index of abstractness, and (3) number of words of personal reference (pronouns, names, words directly relating to people, such as aunt, baby, etc.)(Flesch, 1943b). His regression weights were subsequently recomputed by Lorge when computational errors were found by Dale (Lorge, 1948).

Almost simultaneously, due to difficulty of application, apparent lack of sensitivity to the human interest factor, and misuse of the formula as a rule for writing, Flesch completely revised his formula and published his "reading ease" and "human interest" formulas (Flesch, 1948). The two new formulas were again standardized against the McCall-Crabbs lessons. The "reading ease" formula appeared as R.E. = 206.835 - .846 WL - 1.015 SL, in which SL is average sample sentence length in words and WL is word length measured as syllables per 100 words. Syllable count was substituted for the earlier count of affixes. The correlation between the two measures was .87, and the syllable count was expected to be more easily and reliably taken.

The "human interest" index is presented as:

$$H.I. = 3.635PW + .314PS$$

where PW is average percentage of personal words, using a slightly

narrower definition than previously, and PS is the percentage of sentences spoken or addressed to the reader, including exclamations, or grammatically incomplete sentences whose meaning must be determined from the context. This factor tests the "conversational quality" of the passage, and its inclusion represents an attempt to bring out the easy readability of direct conversational style. The reading ease formula correlated .70 with the McCall-Crabbs; the human interest index correlated .43 with the McCall-Crabbs.

The new formulas were held by Flesch to locate tested passages on scales which range from 0 to 100. A reading ease score of zero is considered to represent "practically unreadable" material, while 100 represents text which is easily read by any literate person. A human interest score of zero indicates no human interest, and 100 indicates that the passage is "full of human interest."

The Flesch formulas have become the most widely applied in the entire history of readability research. This wide application is due in part to the ease of computation of his formulas and partly to the wide exposure given to his formulas through a long series of popularized books (e.g., The Art of Plain Talk [Flesch, 1946], The Art of Readable Writing [Flesch, 1949], and How to Test Readability [Flesch, 1951]). These books saw wide circulation in business, governmental, and journalistic circles, where they were employed as rules for writing (Chall, 1958).

A simplification of the Flesch reading ease formula was proposed by Farr, Jenkins, and Patterson (1951), who observed a correlation of -.91 between the average number of syllables per word and the number of one syllable words in a passage. Accordingly, they suggested that the number of one syllable words be used in Flesch's formula so as to permit more rapid testing with no loss of reliability. Farr, Jenkins, and Patterson generated such a formula, and found it to correlate .95 with scores produced by the Flesch formula. They therefore concluded that their formula,

Reading Ease = 1.599 (number of one syllable words)

- 1.015 (mean sentence length in words) - 31.517

should be considered an acceptable substitute for the Flesch formula.

Both Flesch (1952) and Klare (1952) immediately criticized this proposal. Flesch maintained that accuracy would be lost in evaluating very easy or very difficult materials, and Klare suggested that the reliability of counting one syllable words should be lower than that of counting syllables in the manner suggested by Flesch. England, Thomas, and Patterson (1953) experimentally tested these criticisms and were able to discount them completely. The Farr, Jenkins, and Patterson formula has since received a moderate amount of acceptance--considering the rate at which it is reported as being used in applied evaluations of materials. Chall (1958) pointed out the irony of this return to word length as a criterion of difficulty, one of the very factors Flesch considered unsuitable.

Dale and Chall (1948) employed the Flesch formula to evaluate educational materials published by the National Tuberculosis Association in terms of readability for the average adult. They eventually sought to develop their own measure, because of two drawbacks encountered in the use of the Flesch formula. First, they found low between rater reliability for the count of affixes necessary for the Flesch formula. Although Dale and Chall expressed considerable respect for the justification, presented by Flesch (1943b), for using a count of affixes as an index of abstractness, they wondered (since affixes correlated .78 with abstractness) whether or not some other and more manageable correlate of abstractness could be employed. This question is supported by Lorge (1944), who stated that all measures of vocabulary load, including abstractness, are intercorrelated. Second, Dale and Chall considered the use of personal references in the Flesch human interest formula to be oversimplified. References to senators, though personal, in the Flesch sense, are not generally associated with a lowering of abstractness of text, as are references to "Dad" or "John," which are generally associated with a very concrete type of statement.

Again using the McCall-Crabbs test data originally collected by Lorge, Dale and Chall derived a new regression equation. This equation is based on average sentence length, and the "Dale score" --the relative number of words in the text samples not appearing on a list of 3000 words known to 80 per cent of a sample of fourth graders. The form of the equation was:

$$X_{C50} = .1579 \text{ (Dale score)} + .0496 \text{ (sentence length)} + 3.5365$$

in which the criterion, $X_{C50}$, is the reading grade level of an individual

able to answer correctly half of the comprehension test questions. The score yielded by this equation correlated .70 with the McCall-Crabbs criterion.

Another formula based on the Flesch formula is that of Gunning (1952). In this formula, reading grade level necessary to understand tested material is equal to .4 of the sum of mean sentence length in words and percentage of words of three or more syllables. No numerical correlation was reported between this and other estimates of reading difficulty, although there is little reason to believe that the correlation was low.

The McCall-Crabbs Standard Test Lessons in Reading, which were the basis of the last five formulas discussed, were developed in 1926. They were revised in 1950, and at least 60 of the passages in this more recent edition are entirely new, dealing with modern topics, such as atomic energy and aviation (Powers, Sumner, & Kearl, 1958). Based on this revision of the McCall-Crabbs tests, Powers, Sumner, and Kearl undertook to recompute the Flesch, Dale-Chall, Farr-Jenkins-Patterson, and Gunning indices of readability. They hoped to produce formulas accounting for changes in reading abilities over the twenty-four year period between test editions, and to facilitate comparison of the formulas, since they would be based on identical measurement rules, mathematical procedures, etc.

The recomputed formulas and their respective correlations with the constant criterion of grade score of pupils answering one-half of the test questions correctly are:

Flesch:

-2.2029 + (.0778)(mean sentence length) + (.0455)(number of syllables per 100 words) r = .6351

Dale-Chall:

3.2672 + (.0596)(mean sentence length) + (.1155)(percentage of words not appearing on Dale list of 3000) r = .7135

Farr-Jenkins-Patterson:

8.4355 + (.0923)(mean sentence length)
- (.0648)(percentage of one syllable words) r = .5836

Gunning:

(.0984)(percentage of words of three or more syllables) r = .5865

In support of their recalculations, Powers, Sumner, and Kearl pointed out that the revised formulas give estimates more consistent with one another than did the original Flesch and Dale-Chall formulas; the other formulas were not compared in their original forms. It was concluded that the Dale-Chall formula is the "best," since it had the highest predictive power (correlation) and the lowest standard error, .77 grade levels, of the four formulas tested.

The final readability measure of this period to be discussed is that of McElroy. This measure was unfortunately called a "fog count," the same term applied by Gunning to his formula. This duplication of names has caused confusion on the part of some later researchers, such as Kincaid (1972).

McElroy's fog count is, again, related to the Flesch formula in that it is based upon a count of syllables (Klare, 1963). The procedure to be followed in using this formula is assign a value of 1 to each word of one or two syllables appearing in the sampled passage, and a value of 3 to each remaining word--which will have three or more syllables. The assigned values are added and the value thus determined is the fog count. To determine the reading grade level associated with a particular fog count value, if the sum is over 20, it is divided by 2. If the sum is under 20, 2 is subtracted and the result is divided by 2.

No statistical information relating to the development, or accuracy of this formula is available (Klare, 1963; Kincaid, 1972).

## Recent Formulas

Klare (1963) considers the period 1953 to 1959, when his review was written, to be one exemplified by a trend toward specialization in readability formulas. However, it seems that there is little to be gained in naming a trend on the basis of only seven formulas of relatively minor importance which appeared in a six year period. Extension of the period to include all recent studies of readability measurement techniques appearing from 1953 to the present does, however, appear appropriate and renders impossible the naming of a trend of study characterizing this recent period.

During this period, in addition to formulas intended for a limited range of application, additional readability measures intended for general application were developed, and two new approaches to the problem of measuring readability were presented.

During the last 18 years, four formulas appeared which are intended for application to primary school texts. These are the formulas of Spache (1953), Wheeler and Smith (1954), Bloomer (1959), and Tribe (1956).

The formula of Spache (1953) predicts the primary grade level (grades 1-3) of textual material on the basis of average sentence length in words ($x_1$) and percentage of words not appearing on the Dale list of 769 common words ($x_2$). The formula predicts grade level as equal to: $.141x_1 + .086x_2 + .839$. The reported correlation between predicted score of tested materials and usual grade level of application was .818.

Wheeler and Smith (1954) based their formula for prediction of grade level on the grade designation assigned by the publisher of the textual materials in their sample. Their formula equated grade levels to 10 times the product of the mean length of units (sentences, with minor exceptions) in words and the percentage of multisyllable words. The value thus determined is located in a table and grade level of 1 through 4 read off. This was the first formula to be based on a multiplicative model of factors. This type of equation allows interactions between the factors (as discussed by McLaughlin [1969]), so that a particular change in factor A will affect the predicted value differently at varying levels of factor B. This may be a highly important

characteristic for a readability formula, in the light of data such as that of Gray and Leary (1935), showing that the relative importance of vocabulary and structural factors does not remain constant across all reading levels. Multiplicative formulas may be able to deal more appropriately with findings of this type than do additive formulas. But, it is doubtful that the simplified formula of Wheeler and Smith is a very large step in the proper direction.

The formula of Bloomer (1959) again used customary grade level application as the readability criterion. In Bloomer's development, reading grade level is predicted from abstraction level as indicated by the number of words per modifier (modifying phrase) and "sound complexity" (sic) of modifiers. Bloomer contended that these variables may be used as predictors of readability, although he presented no formula, predictive method, or method for application. His approach is based on the observation that abstraction in text increases with grade level, and that the two variables (words per modifier and sound complexity) employed are closely associated with abstraction. The multiple correlation between the two variables and assigned grade level was .78, which Bloomer considers to compare favorably with the correlations obtained through the procedures of Flesch and of Lorge.

The McCall-Crabbs Standard Test Lessons in Reading, 1950 revision, were again used as a criterion by Tribe (1956). Grade level score of children, in grades 2 through 8, who could correctly answer one-half of the reading test questions was found to be equal to:

$$.0719x_1 + .1043x_5 + 2.9347$$

where $x_1$ is the average sentence length and $x_5$ is the percentage (times 100) of words not appearing on the Rinsland word list. A correction factor is then applied to the predicted grade level. This factor is based on a table presented by Dale and Chall (1948).

An unusual criterion was employed by Jacobson (1965) to develop formulas to determine the readability levels of high school and college chemistry and physics texts. The predicted criterion measure was the average number of words indicated as not being understood (as indicated by reader underlining) by readers, based on 200 word sample passages. This procedure was first employed in 1928, and is known as the Kyte test. Test-retest reliability of this test over a one week interim was reported by Jacobson to be .95 for the physics texts and .85 for the chemistry texts in his sample. Two formulas are presented--one for physics texts and one for chemistry texts. The variables included are:

$X$ mean underlining score (the criterion)

$x_1$ mean number of words per independent clause in each 200 word sample passage

$x_2$ mean number of mathematical terms in the sampled passages

$x_3$ mean number of words in the sampled passages that are above the 6,000 word level in Thorndike's 20,000 word list

$x_4$ mean number of (technical) words per passage which do not appear in the Powers list of 1828 essential scientific terms

The formula derived for use with physics texts is:

$$X = -.0003 + 29.7059x_2 + 21.119x_3 + 35.0029x_4$$

which exhibited a correlation with the criterion of .70.

The formula for chemistry texts is:

$$X = .003 + .1706x_1 + 13.7231x_2 - 43.7262x_3 - 2.3577x_4$$

Predictions from this formula correlated .67 with the criterion.

Flesch introduced three additional readability indexes (the term "formula" is probably inappropriate here) in 1954 and 1958 (Flesch, 1954, 1958). All are relatively subjective measures in which counts are made, based on 100 word samples of text, and the counted values are converted to arbitrary scales through reference to conversion tables. All were validated by inspection only.

The "r" score (Flesch, 1954) is an index of realism, based on the number of references to specific human beings, their attributes or possessions, locations, objects numbered or named, dates, times, and colors.

The "e" score (Flesch, 1954) is an index of energy, based on indications of voice communication, such as inflection.

The "formality-popularity" scale (Flesch, 1958) is based on the total numbers of: capitalized, underlined or italicized words, numbers (not spelled out), punctuation marks, symbols (#, $, ¢, etc.), beginnings of paragraphs, and endings of paragraphs.

Forbes developed a readability measure intended for application to the instructions and items contained in all types of standardized tests and opinion polls, except vocabulary tests (Forbes & Cottle, 1953). To determine a test's reading grade level, each word appearing difficult to the grader is looked up in the 1942 Thorndike Junior Century Dictionary, and its listed frequency of occurrence, from the first to the 20,000th words noted. The total of all indices above 4 is computed and divided by the total number of words sampled. This vocabulary index is looked up in a table, which indicates the corresponding grade level.

They reported that the readability of tests measured using this procedure correlates .96 with the average of the readability of the same material calculated using five other procedures, including the Dale-Chall and Flesch methods. According to Forbes and Cottle, at the time of this work, no readability measure directly applicable to test forms was available. However, in view of the reported high correlation (.96) between this technique and the others, it seems that the contribution is minimal. Additionally, if the other measurement procedures are not appropriate to this application, what is the value of the Forbes and Cottle contribution? No other method of assessing the readability of the test materials is reported.

Six new readability measures of the traditional form have appeared in very recent years. The first of these was that of Coleman, developed in 1965 and discussed by Szalay (1965). Coleman developed a family of four formulas, using one through four measured variables, respectively, to predict readability level. The readability criterion in this case was the mean "cloze" score on the passage achieved by a sample of college students. (The "cloze" score is the percentage of deleted words of a passage that are correctly guessed and written in by a subject. This approach to readability is discussed in subsequent paragraphs.)

Correlations among the four formulas and criterion scores varied between .85 and .91 when tested independently by Szalay and Coleman (Szalay, 1965). It is therefore recommended by Szalay that only the simplest formula be employed:

$$\text{Predicted cloze score} = 1.29\,(\text{percentage of one syllable words}) - 38.45.$$

Factors present in the other three formulas but not making significant additions to predictive ability are: (1) sentence length, (2) frequency of occurrence of pronouns, and (3) frequency of occurrence of prepositions.

A very similar formula was developed by HuMRRO personnel for measuring the readability of Army technical literature (Caylor, Sticht, Fox, & Ford, 1972). Their measure is called the FORCAST formula after its developers, FORd, CAylor, and STicht. They considered existing formulas inappropriate for their purpose because school students and school or general texts had been most often employed in developing readability formulas. This type of standardization was believed to make the applicability of prior formulas to technical publications for adults suspect. Moreover, application of many of the existing formulas required special grammatical or linguistic competence on the part of the person attempting to apply the formulas. Ford, Caylor, and Sticht elected to use cloze score as the criterion of readability. They believed the cloze test to be more objective than multiple choice tests or the other more traditional indices of comprehension. They also pointed out that cloze has "consistently yielded very high correlations with multiple choice tests and other more subjectively constructed measures of comprehension and difficulty" (Caylor, et al., 1972, p. 12).

Additionally, as part of their own work, they found a correlation of approximately .80 between cloze score on 150-word passages chosen from the readings required in a wide range of Army jobs and achieved reading grade level, as measured by the United States Armed Forces Institute Reading Achievement Test III, Form A, Abbreviated Edition.

Previous research had indicated that if a cloze score of 35 per cent is achieved by a given subject on a particular test passage, then it may be reasonably expected that the subject will correctly answer approximately 70 per cent of a set of multiple-choice questions based on that passage. Hence, cloze score appears to be a good indicator of both comprehension and reading achievement level.

A cloze score of 35 per cent was arbitrarily chosen as the criterion of potentially adequate comprehension of a text passage. The reading grade level of a passage was defined as the lowest reading grade level (as determined by USAFI Reading Achievement Test score) at which 50 per cent of the tested subjects achieved a cloze score of 35 per cent or higher on that passage.

A literature search provided Ford, Caylor, and Sticht with a list of 15 structural properties of text that had been applied in previous readability formulas and required no special competence or equipment to measure. Correlations between cloze score and each of the structural properties were computed and several regression equations were derived. Their preferred formula employed only a single factor, number of one-syllable words per passage. This factor is very easily measured, and basing the equation on additional factors allowed no practical increase in predictive power.

The FORCAST formula predicts reading grade level as equal to:

20 - (number of one-syllable words/10)

The correlation between predicted reading grade level of a passage with tested reading grade level associated with 35 per cent cloze score was .87. A subsequent application using new test passages and new subjects produced a correlation of .77.

The two last discussed formulas are extremely simple in form. However, the inclusion of additional factors in the formulas allowed inconsequential gains in predictive power, while greatly increasing the effort required in applying the formulas. In contrast, the developers of the remaining four readability measures of recent years set out with the stated purpose of attempting to provide a method of measuring readability with minimal effort.

Smith and Senter (1966) developed a readability equation whose data may be collected from mechanical counters easily installed on an IBM Selectric typewriter. This technique allows measurement of readability at essentially rough draft typing speed. Mechanical counters are used to record the numbers of key strokes, blank spaces, and sentences (an equal sign must be typed at the end of each sentence; the number of activations of this key indicating the number of sentences typed). From these counts, the mean number of words per sentence [number of spaces divided by number of sentences (w/s)] and the mean length of words (number of strokes divided by number of spaces (s/w)] may be computed. Based on examination of graded school texts, the regression equation predicting grade level (GL) from the above ratios is:

$$GL = 0.50\ (w/s) + 4.71\ (s/w) - 21.43.$$

This may be simplified to yield the arbitrarily scaled Automated Readability Index (ARI) equal to (w/s)+ 9 (s/w). The authors support use of the ARI instead of predicted grade level because considerable variability exists in characteristics of texts written for school students beyond the junion high school level. Accordingly, a precise statement of grade equivalent appears inappropriate. It is also pointed out that readability, as measured by a formula such as this, increases more slowly with grade level at high levels than at low ones.

Dismissal of prediction of grade level removes the need to complicate the formula by attempting to deal with this nonlinearity.

As advantages of this procedure of estimating readability, Smith and Kincaid (1970) pointed out: (1) the speed of data collection that is possible, (2) the concrete nature of the acquired data, making its collection extremely objective and reliable, and (3) the ease with which it could be incorporated into modern computerized typesetting machinery.

Coke and Rothkopf (1970) have adapted the Flesch readability formula for automatic computation by computer. The determination of words per sentence is straightforward in their algorithm, but word length in syllables is indexed by number of vowels per word. Using these two variables, their program produced Flesch reading ease scores that exhibited a correlation of .92 with scores calculated using the normal procedure. They discount the practical utility of their program, but indicate that it is useful for testing the adequacy of text sampling procedures. They present a graph showing the probability of miscalculating reading ease score by five points or more as a function of sample size.

Fry (1968) presented his readability measure as a simple graph on which grade level may be looked up, given average sentence length in words and syllables. With this presentation, he hoped to reduce greatly the amount of time required to compute an index of readability, thereby increasing the popularity of such a measure. His grade level designations were based on inspection of textbooks used at various grade levels. The graphic presentation employed permits the readability measure to reflect accurately nonlinearities in the grade level function without resorting to higher order equations or arbitrary scaling.

McLaughlin, a psycholinguist, developed a readability formula which is based on the 1961 revision of the McCall-Crabbs Test Lessons. His formula is based on a multiplicative rather than additive model of factors (McLaughlin, 1969). McLaughlin feels that word length, a measure of semantic difficulty, and sentence length, a measure of syntactic difficulty, interact with reading difficulty in a manner that cannot be accounted for by additive formulas. Interestingly enough, McLaughlin found that his readability formula, based on a multiplicative model, could be computed more easily than any previously existing measure. He first points out that the step of multiplication of word and sentence lengths may be avoided, since a count of syllables in a set number of sentences is an equivalent procedure. He next points out that even counting all the syllables in N sentences is unnecessary. He found that the number of syllables in 100 words is equal to three times the number of words of over two syllables, plus 112. He then derived his regression equation, and found that by adjusting tested sample size the formula for predicting grade level necessary to answer 100 per cent of the McCall-Crabbs questions correctly could be greatly simplified. A very close approximation to his formula is presented:

$$\text{SMOG Grade} = 3 + \sqrt{\text{polysyllable count}^{*} \text{ in 30 sentences}}$$

Predictions from this formula correlated at approximately .70 with the criterion. The term SMOG grade, or SMOG count, is in tribute to the FOG count of Gunning, the first application of the count of number of polysyllabic words to readability determination and to the characteristic atmospheric condition of his home city--London.

---

*number of words of three or more syllables.

## Cloze Procedures

The cloze procedure, which has been briefly mentioned previously, was introduced by Taylor (1953) as a measure of readability that is free from many of the disadvantages of traditional readability measures. In the cloze method, samples of text are presented with some words deleted and replaced by blank spaces. The subject's task is to fill in the blank spaces with the correct words. The name "cloze" was applied to this procedure by Taylor because of its resemblance to the principle of closure of the Gestalt school of psychology: the "human tendency to complete a familiar but not-quite finished pattern--to 'see' a broken circle as a whole one...by mentally closing the gaps."

In his initial report, Taylor presented data showing that the cloze procedure consistently ranked tested "standard" reading passages in the same order as the readability formulas of Flesch and of Dale-Chall. He also indicated that the rank ordering of cloze scores is maintained regardless of system of word deletion employed, be it every nth word or random, with low (10 per cent) or high (20 per cent) rates of word deletion. A random or, equivalently, every nth deletion system is strongly defended. If enough words are deleted, all kinds of words are deleted in the proportion in which they actually occur in the text.

Analysis of the effects of scoring for only precise matches with the deleted word as compared with applying the more tedious procedure of accepting synonyms as correct responses raises all scores equally. Accordingly, there is no effect on discriminability and the more difficult procedure is not warranted.

Taylor also demonstrated that the cloze test can handle unusual materials more effectively than readability formulas. Gertrude Stein, for example, writes in quite short sentences with a fairly simple vocabulary. However, her style is such that her material is very difficult to read. This is accurately reflected by the cloze test, but the Flesch and Dale-Chall formulas rate sample passages taken from her writings as very easy reading.

Taylor (1957) found correlations of .70 to .80 between cloze scores and comprehension scores of Air Force trainees reading typical Air Force technical material. Bormuth (1968) found correlations of .90 to .96 between cloze scores and scores on tests of comprehension of passages from the Gray Oral Reading Tests. Bormuth's questions were of the transformational type, measuring retention of "facts" from the passages only. In constructing questions of this type, one word or clause of a statement in the passage is deleted and replaced by a question marker. The answer to the question, then, is the deleted element. For example, "The boy rode the horse," becomes "Who rode the horse?" Bormuth indicated that limiting the test to questions of this type circumvented the problem of poor matching of the readability levels of the passages and the tests, since the questions are determined by the sentences of the passage.

Rankin and Culhane (1969) similarly correlated cloze scores and comprehension scores, but without limiting the types of questions included in the comprehension tests. Their test questions included items relating to vocabulary, fact, sequence, causal relationship, main idea, inference from facts, and author's purpose. They obtained a correlation of .68 between cloze scores based on excerpts from encyclopedia articles and comprehension test score.

Bormuth (1967) determined the cloze scores corresponding to: (1) 75 per cent comprehension test score, the comprehension level generally considered necessary to allow effective classroom study of a text with assistance of a teacher available, and (2) 90 per cent comprehension test score, the level which is considered an indication that the text is sufficiently comprehensible to allow effective independent study. A 30 per cent cloze score was found to be associated with 75 per cent comprehension test score, and 50 per cent cloze score was associated with 90 per cent comprehension score. Replication of the study in the succeeding year (Bormuth, 1968) showed a cloze score of 44 per cent to be associated with the "classroom level" and 57 per cent to be associated with the "independent level." Similar procedures by Rankin and Culhane (1969) using their more difficult type of comprehension test, as described previously, found cloze scores of 41 and 61 per cent at the comprehension score points of interest. The correspondence between these results and those of Bormuth is quite remarkable in view of the fact that Bormuth considered his 1967 results

relating cloze score to 90 per cent comprehension score relatively invalid, because of ceiling effects present in the comprehension test used in that study.

Based on these results, Rankin and Culhane concluded that it would be appropriate for teachers to consider books on which pupils cannot attain a cloze score of approximately 40 per cent to be too difficult for those students.

The cloze procedure has numerous advantages and disadvantages. Among the advantages pointed out by Taylor (1953) and Klare, Sinaiko, and Stolurow (1970) are: (1) scoring reliability is very high, (2) it works well with "non-standard" material, (3) it accounts for interest and prior knowledge of reader populations, (4) subjects of all abilities seem to enjoy it, and (5) test materials are easy to construct.

Among its disadvantages are: (1) cloze is a measure, not a predictor of readability, requiring testing of sizable samples of people, (2) Klare et al. (1970) hypothesize that it may depend more on knowledge of language than subject matter, (3) it may not accurately reflect all types of comprehension, and (4) it may depend excessively on "short-range constraints"--the four or five words appearing on each side of the deleted word.

## Application of Readability Formulas as "Rules for Writing"

The urge to apply mechanically the traditional readability formulas for purposes of improving readability is very strong, but may not be appropriate. Smith and Kincaid (1970) point out that readability scores are gross measures of difficulty at best and must not be taken as indices of good or bad writing. A deliberate attempt to shorten sentence and word length does not necessarily enhance readability. In fact, readability may be degraded. A more advantageous approach is to make the writing more logical and precise. When combined with these principles, consideration of the structural factors contained in readability formulas may, however, contribute to readability.

The recommendations of Flesch (1951) are typical of those made in hope of improving readability. He suggests that: (1) a personal type of discourse be adopted, (2) the importance of points presented be discussed, (3) introductions and summary statement be included, (4) punctuation be used in such a manner as to help the reader, (4) points be discussed in chronological order or in order of increasing importance, and (6) excessive wordiness be avoided. He does recommend, additionally, that short paragraphs, sentences, and words be used.

The advantage to be gained by careful consideration of the overall structure of a passage of text was demonstrated by Lee (1965). Significantly better learning was found from articles written in a highly structured manner compared to "normal" articles or those in which paragraph order has been randomized. The highly structured articles differed from the normal in that they included: (1) an introductory paragraph outlining the points to follow, (2) a final summarizing paragraph, (3) a number of major and minor headings, and (4) transitional paragraphs elaborating on completed and subsequent topics and emphasizing the organization of the paper.

## Multimodal Presentation

The field of information transfer has also turned to the investigation of multimodal information presentation to enhance and facilitate the information transfer process. Research in this area has been mainly devoted to answering the question: Can method X transfer information as well as some other method? It has been argued (Phillips, 1966) that this is not a worthwhile issue because the question assumes that the method against which comparisons are being made has been the most effective in transmitting the information. What the research may actually demonstrate is that method X may transfer information just as poorly as method Y. Phillips (1966) suggests that a more relevant question is:

> **Which resource or combination of resources (people, places, media) is appropriate for teaching what type of subject matter to what type of learner under what conditions (time, place, size of group, and so on) to achieve what purpose? (p. 374).**

Research on improving instructional materials has often occurred mainly without any reference to any precise theoretical notions (Briggs, 1966; Lumsdaine, 1964). Theories of learning have not been taken into account. Even though a great deal of research has been performed to improve the effectiveness with which materials are presented in certain media, when one asks which media will be more effective in presenting a certain type of material to a special class of learner, one comes to a standstill.

An experiment by Bourisseau, Davis, and Yamamato (1965) demonstrated how assumptions in the audiovisual field often fail to possess merit. It is generally assumed that in terms of direct sensory appeal, pictures are superior to printed or spoken words. "A picture is worth a thousand words" is generally assumed. Contrary to this belief, Bourisseau, Davis, and Yamamato (1965) demonstrated that pictorial stimuli are inferior to verbal (printed) stimuli in regard to both the number of subjects making sensory responses and the total number of sensory responses evoked.

There have been a number of studies which suggest little, or no payoff from multimodal presentation.

Virag (1971) attempted to assess the effectiveness of three modes in transmitting content to students with different aptitudes. The three instructional modes were: (1) low verbal (tape-slides and short film episodes--materials presented at a fixed pace via the audio-visual communicative channels), (2) high verbal (written case studies --presented at a self pace through the channel of print), and (3) conventional (short lectures--presented at a fixed pace through the audio-communicative channel). The results indicated that no single mode of instruction was consistently more effective than the other two for any particular aptitude pattern.

In an attempt to prepare an audio-tutorial minicourse, Long (1970) made use of five methods. The procedures presented the materials to be learned through: (1) printed text, (2) printed text with supplementary programmed items, (3) printed text with supplementary laboratory manipulations and observations, (4) taped audio program with supplementary programmed items, and (5) taped tutorial program with laboratory manipulations and observations. No significant differences were found to exist between the five instructional methods employed.

Travers (1965) reported a study performed by Van Mondfrans in which verbal materials to be learned were presented auditorially, visually, and simultaneous audio-visual presentations. The results indicated that there was no difference between single sense channel presentations and no difference between single and multiple sense channel presentations. This research supports the earlier conclusion by Van Mondfrans and Travers (1964) that the use of two sensory modalities has no advantage over one in the learning of material which is redundant across modalities.

In a study by Goodrich (1971), four types of literary forms or subjects were presented via four instructional media: (1) fiction, (2) nonfiction (autobiography), (3) a lecture about literary symbolism, and (4) a lecture about composition. The four media were: (1) TV, (2) audio tape, (3) face-to-face lecture, and (4) text. There were no significant differences between the different media. Goodrich concluded that the effects of the medium may not be noticeable under normal learning situations.

Sticht (1969) presented materials of differing levels of difficulty to Ss of different aptitudes through the visual and auditory sensory channels. According to Sticht, the results indicated that listening was as effective as reading in transmitting information of all three difficulty levels for both average and low aptitude subjects. Siegel, Barcik, and Macpherson (1965), at Applied Psychological Services, presented materials to four groups of college students via the auditory and visual channels. Two groups, one auditory and one visual, also received adjunct programmed materials. The adjunct materials consisted of multiple choice questions with correct answers. The use of the adjunct materials provided feedback to the learner on what information was missed. The results indicated that both audio and visual presentation with adjunct programmed materials were significantly better than without adjunct materials. There were no significant differences between the different sensory channel presentations.

On the other hand, a number of studies have suggested some gain to accrue from multichannel presentation.

Singer (1970) investigated comprehension as effected by varying visual and auditory presentation. The information to be learned was presented: (1) textually, (2) verbally, and (3) verbally paired with reading. Comprehension was poorer when the materials were presented only auditorially, but no difference existed between reading and reading with listening.

Severin (1967) presented materials via the auditory, the visual, and the audio-visual modes of presentation. The results indicated that print and print with audio were superior to audio for transferring information. There was no difference between print and print with audio. These findings were substantiated in a later study (Singer, 1970).

The results of a recent study (Senour, 1971) concerning the effects of student control of audio tape learning indicated that providing control functions to the student; i. e., the capability for starting, stopping, and replaying the tape, aided learner achievement as compared to the situation which denied them. There was a significantly positive correlation between learner achievement and the number of times the learner elected to use the controls. The subjects reported that they did not have to take notes because they could replay the tape until they knew it.

Nelson (1970) reported a study on the effects of visual-auditory modality preference on learning mode preference. The results indicated that auditory and visual perceptual subtests of reading readiness tests are not sufficiently sensitive to discern modality preferences. Another report (Hueber, 1970) supported the results obtained by Nelson (1970).

If sensory modality preferences do affect information transfer, a means of determining these preferences must be developed. If audio tape presentations of information are going to be used, can subjects be taught to listen more attentively? A large body of research indicated that training to listen is possible (Brown, 1954; Erikson, 1954; Irwin, 1953; Nichols, 1949; Lewis, 1956). Other researchers (Erikson, 1954; Irvin, 1954) suggest that low listening ability subjects benefit from such training more than subjects of high listening ability.

# CHAPTER III

## DISCUSSION

The most important characteristics of the readability formulas described here, as well as of those formulas appearing before 1953, and considered of at least moderate importance are presented in tabular form in Appendix A to this review.

In terms of application of these formulas for predicting readability, in the late fifties, there appeared to be rather general agreement (Chall, 1958; Klare, 1963; Powers et al., 1958), that the most precise formula available, and therefore the one to be most generally recommended, was the Dale-Chall method. If reference to a word list was to be avoided, the Flesch formula was recommended, unless other factors such as special reader populations, types of reading material, or particular advantages of one or another formula warranted another choice. It does not seem that this general set of recommendations should be changed at this time.

### Estimation of Reading Level

The majority of the readability formulas predict readability in terms of reading grade level. In order to apply them most effectively, knowledge must be obtained concerning the distribution of reading grade levels of the expected reader population. The variability of reading grade level within school grades and within groups of adults having achieved particular levels of education found by Gray and Leary (1935) suggests that pure estimation of reading grade level is somewhat unreliable. Reading grade level may be determined by administering standard tests of reading ability, of which 37 are reported as being available in The Sixth Mental Measurements Yearbook (Buros, 1965), to a sample of the expected reading population.

However, in certain areas of application, most notably the military, a much more efficient, less time consuming, and less expensive procedure may be employed. It has been found in the Air Force (Madden & Tupes, 1966) and in the Army (Caylor et al., 1972) that reading grade level may be estimated from certain aptitude test scores. Madden and Tupes noted that the general aptitude index (AI) of the Airman Qualifying Exam (AQE) correlated above .70 with reading level. This is largely due to the inclusion of a reading vocabulary subtest score within the general AI. Although reading grade level as measured by the California Test of Reading Vocabulary and Reading Comprehension was estimable from the general AI alone, more accurate prediction of an individual's reading grade level could be made by using a regression equation based on general AI and the individual's selector AI score. The latter is one of three aptitude area scores--administrative, mechanical, or electronic--which are referred to when assigning men to career fields in the Air Force. For some career fields, the selector variable is the administrative AI; for others, it is the mechanical AI, while for others, the electronic AI is employed. Selection to a few career fields is based solely on general AI. The regression equations appropriate for estimating reading grade level of individuals in career fields which use the selector AIs for personnel selection are:

administrative: RGL = .0437(GenAI) + .0501(AdAI) + 5.0730

mechanical: RGL = .0991(GenAI) + .0085(MechAI) + 5.0459

electronic: RGL = .0743(GenAI) + .0222(ElAI) + 4.6088

Caylor et al. (1972) similarly produced a regression equation to predict reading grade level as measured by the U. S. Armed Forces Institute (USAFI) Reading Achievement Test III, Form A, from knowledge of an individual's Armed Forces Qualifying Test (AFQT) score. Their equation is:

RGL = .75(AFQT score) + 5.52.

These regression formulas allow much more confidence to be placed in statements concerning the appropriate readability levels of materials intended for use by Army or Air Force enlisted personnel.

## Discussion of Formulas

A number of general criticisms have been applied to readability formulas. The definition of the criterion of comprehensibility has been a problem since the earliest studies (Bormuth, 1966). The usual practice has been to administer multiple choice criterion questions just after the passages being tested are read. Lorge (1939) criticized this procedure because test performance may be strongly influenced by the difficulty of the language of the test questions.

The difficulty of the questions may also be varied, so that the subject may be able to score highly by simply remembering details of a passage, or he may be required to determine the author's purpose in writing the passage, and select the best of several highly similar alternatives.

In addition, Fry (1968) pointed out that reading grade levels are not rigorously defined, so that different reading tests, especially those developed at different times, may provide different grade levels for identical subjects. New reading tests are more difficult than old ones, indicating that students at given grade levels read better than their predecessors. He summarizes the problem as one of "trying to determine grade level when grade level won't stand still." It seems strange that the various predictions have not been corrected for criterion unreliability. Also, there is no agreement on the level of comprehension to be accepted. Fifty and 75 per cent were common when McCall-Crabbs tests were employed, but the FORCAST formula uses 70 per cent (35 per cent cloze score) and the SMOG count was based on 100 per cent comprehension.

Most of the readability formulas do not account for the effects on readability of the reader's interests, experiences, or aptitudes. Exceptions to this are the supplementary indices of human interest, abstraction, realism, energy, and formality-popularity of Flesch,

along with the FORCAST formula, and Jacobson's measures of the readability of physics and chemistry texts. However, some of these measures will allow accurate reflection of the readability of material appearing in professional or technical journals.

Finally, as Bormuth (1966) pointed out, until very recent years no theoretical base was available from which to generate testable hypotheses relating to readability. Powerful theories of language behavior did not exist, so that only the most obvious statistical characteristics of the written text were studied.

Although he did not consider it appropriate to present his formulas in his 1966 paper, Bormuth reported that he has written regression equations whose predictions correlate up to .93 with cloze test score. He considers cloze scores to be the only acceptable criterion of readability currently available. He reported that all of the variables in his formulas are new ones generated from modern linguistic theories such as those of Chomsky and Yngve. None of the traditional readability variables were powerful enough to be included in his formulas.

Inter-user and test-retest reliabilities of formulas vary widely. Kincaid (1972) found that the test-retest reliability of the McElroy fog count was only .5. He also pointed out that he gets a headache after taking a fog count for 30 minutes. The highly objective ARI, however, has a measured test-retest reliability of over .99 (Huff, 1970). England, Thomas, and Patterson (1948) found the interrater and test-retest reliabilities of the Flesch reading ease formula were approximately .90.

The reliabilities of other readability formulas had not been satisfactorily tested prior to 1958, according to Klare (1963), and work of this type has not appeared since that time.

All of the readability measuring procedures discussed show roughly similar correlations with their respective criteria of difficulty. Additionally, all of the criteria seem to be highly intercorrelated. Hence, from the validity standpoint, there is little basis for selecting one readability measure over another. In this case, utility and practicality for the individual user seem to represent the criteria to be employed when selecting a readability measure. As has been previously reported, the Dale-Chall and Flesch formulas, as revised by Powers, Sumner, and Kearl (1958), are the most highly respected of the traditional formulas, but considerations of particular characteristics of an individual situation may warrant choice of one of the many other available formulas.

There is, however, some indication that the most powerful method of measuring readability currently available is the cloze method. However, cloze is less easy to apply than the other techniques. Application of the cloze method requires preparation of numerous test forms, assembly of a group of subjects who are representative of the appropriate reader population, and considerable administration and scoring time. Also, cloze tests cannot be used to monitor the readability of material as it is being written.

It has been known, since at least the time of the Dale-Tyler study in 1934, that using readability formulas as a basis for "rules for writing" is not an effective way to produce readable material. The recommendations made to writers interested in improving the information transfer capability of their output have always been of a "qualitative" nature, as opposed to the "quantitative" nature of the readability formulas. Recommendations to writers concern such considerations as stylistic variables, level of abstractness, coherency, obscurity of expression, difficulty of ideas expressed, ideational density, and so on. These factors are generally ignored by readability formulas, since they are extremely difficult to measure or even define. This leaves us in the paradoxical situation of attempting to measure readability by measuring factors which do not determine reading ease. Readability formulas are based on relatively unimportant characteristics of text which must be considered almost entirely artifactual.

## Future Research Avenues

The search for more relevant variables and methods of measuring them impresses us as the most pressing need for future research in the area of readability measurement. It is likely that modern psycholinguistic theory could be a very stimulating area from which to draw important readability variables. Applications of information theory to the study of readability have not proven fruitful thus far. Such applications have been limited to attempts to measure the redundancy of textual material as reflected by the variability of responses made in cloze tests.

In order to measure information transfer using the concepts of entropy and bits, it is necessary to be able to specify accurately the stimulus alphabet. In this case, the stimulus variables are letters or words and the probabilities and conditional probabilities of occurrence of each symbol, from the point of view of the receiver, or reader. Inability to specify adequately these probabilities makes the information theoretic approach seem relatively inappropriate at this time.

Research is also needed which focuses on development of a manageable criterion of readability. Cloze scores do not represent a panacea, but the other criteria available have disadvantages. Comprehension test scores are easily confounded with variable aspects of the test questions, and reading grade level is another step removed from comprehension, the topic of interest.

Second, the effects of the additional variables of general intelligence, interest, and prior experience on readability are not well understood. In 1935, Gray and Leary observed that quantitative factors related to readability were not the same for good and poor readers. But work on this variable has not continued. Few, if any, investigations have sought to determine the effects of reader interest and experience on readability, although a few formulas have attempted to account for these factors, notably the Flesch reading ease formula and the FORCAST formula.

Third, the majority of readability formulas have been validated on school students of various levels. Students comprise the population of interest in only a small portion of the possible applications of readability formulas. Hence, more cross-validational studies are needed which are based on samples of adult readers of all reading ability levels.

Fourth, no measure of readability is available for evaluation of tests or programmed instructional material. These types of material are being used more and more in our society, and a method of measuring their readability is greatly needed.

In all future research, it will be appropriate to evaluate each readability formula developed in mathematical forms other than the additive linear model traditionally employed. Many authors have found nonlinearities in the relationships between their chosen variables and criteria. They have dealt with this problem by presenting their data graphically, or presenting a table of corrected values for the predictions from their formulas. It is likely that nonlinear regression techniques account for available data in a significantly more adequate manner than do formulas of the current linear variety.

## REFERENCES

Bloomer, R.H. Level of abstraction as a function of modifier load. Journal of Educational Research, 1959, 52, 269-272.

Bormuth, J.R. Readability, a new approach. Reading Research Quarterly, 1966, 1, 79-132.

Bormuth, J.R. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1967, 10, 291-299.

Bormuth, J.R. Cloze test readability: criterion referenced scores. Journal of Educational Measurement, 1968, 5, 189-196.

Bourisseau, W., Davis, O., & Yamamato, K. Sense-impression responses to differing pictorial and verbal stimuli. AV Communication Review, 1965, 13, 249-258.

Briggs, L. A procedure for the design of multimedia instruction. In Proceedings of the 74th Annual Convention of the American Psychological Association. Washington, D.C.: American Psychological Association, 1966. P. 257-258.

Brown, J. How teachable is listening? Education Research Bulletin, 1954, 33, 85-93.

Buros, O.K. (Ed.). The Sixth Mental Measurement Yearbook. Highland Park, N.J.: Gryphon Press, 1965.

Caylor, J.S., Sticht, T.G., Fox, L.C., & Ford, J.P. Methodologies for determining reading requirements of military occupational specialities. Presidio of Monterey, Calif.: Human Resources Research Organization, HumRRO Division No. 3, Draft Technical Report, January 1972.

Chall, J.S. Readability: An appraisal of research and application. Columbus, Ohio: Bureau of Educational Research, Ohio State University, 1958.

Coke, E.U., & Rothkopf, E.Z. Note on a simple algorithm for a computer-produced reading ease score. Journal of Applied Psychology, 1970, 54, 208-210.

Dale, E., Chall, J.S. A formula for predicting readability (and) Instructions. Educational Research Bulletin, 1948, 27, 11-20, 28, 37-54.

Dale, E., & Chall, J.S. The concept of readability. Elementary English, 1949, 26, 19-26.

Dale, E., & Tyler, R.W. A study of the factors influencing the difficulty of reading materials for adults of limited reading ability. Library Quarterly, 1934, 4, 384-412.

De Long, V.R. Primary promotion by reading levels. Elementary School Journal, 1939, 38, 663-671.

Dolch, E.W. Fact burden and reading difficulty. Elementary English Review, 1939, 16, 135-138.

Dolch, E.W. The use of vocabulary lists in predicting readability and in developing reading materials. Elementary English, 1949, 26, 142-149.

England, G.W., Thomas, M., Paterson, D.G. Reliability of the original and the simplified Flesch reading ease formulas. Journal of Applied Psychology, 1953, 37, 111-113.

Erikson, G. Can listening efficiency be improved? Journal of Communication, 1954, 4, 128-133.

Farr, J.N., Jenkins, J.J., Paterson, D.G. Simplification of Flesch reading ease formula. Journal of Applied Psychology, 1951, 35, 333-337.

Flesch, R. Estimating the comprehension difficulty of magazine articles. Journal of General Psychology, 1943, 28, 63-80. (a)

Flesch, R. Marks of readable style. New York: Bureau of Publications, Teachers College, Columbia University, 1943. (b)

Flesch, R. The art of plain talk. New York: Harper & Bros., 1946.

Flesch, R. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.

Flesch, R. The art of readable writing. New York: Harper & Bros., 1949.

Flesch. R. How to test readability. New York: Harper & Bros., 1951.

Flesch, R. F. Reply to "Simplification of Flesch reading ease formula." Journal of Applied Psychology, 1952, 36, 54-55.

Flesch, R. F. How to make sense. New York: Harper & Bros., 1954.

Forbes, F.W., & Cottle, W.C. A new method for determining readability of standardized tests. Journal of Applied Psychology, 1953, 37, 185-190.

Fry, E. A readability formula that saves time. Journal of Reading, 1968, 11, 513-516, 575-578.

Gillie, P.J. A simplified formula for measuring abstraction in writing. Journal of Applied Psychology, 1957, 41, 214-217.

Goodrich, H. An investigation of the differential effects of four different media on information acquisition and preception. Dissertation Abstracts International, 1971, 31, 3775.

Gray, W.S., & Leary, B.E. What makes a book readable. Chicago: University of Chicago Press, 1935.

Gunning, R. The technique of clear writing. New York: McGraw-Hill, 1952.

Huebner, R. Interaction between patterns of individual differences in sensory modalities of students and methods of classroom instruction. Dissertation Abstract International, 1970, 31, 679.

Huff, K.H. & Smith, E.A. Reliability, baseline data, and instructions for the automated readability index. AFHRL-TR-70-14, AD-729 209. Lowry AFB, Colo.: Technical Training Division, Air Force Human Resources Laboratory, October 1970.

Irvin, C. An analysis of certain aspects of listening training program among college freshman at Michigan State College. Journal of Communication, 1954, 4, 2.

Irwin, C. Evaluating and training program in listening for college freshmen. School Review, 1953, 61, 25-29.

Jacobson, M. D. Reading difficulty of physics and chemistry textbooks. Educational and Psychological Measurement, 1965, 25, 449-457.

Johnson, G. R. An objective measure of determining reading difficulty. Journal of Educational Research, 1930, 21, 283-287.

Kessler, E. The readability of selected contemporary books for leisure reading in high school biology. Science Education, 1941, 25, 260-264.

Kincaid, J. P. Making technical writing readable. Human Factors Society Bulletin, 1972, 15(5), 3-4.

Klare, G. R. A table for rapid determination of Dale-Chall readability scores. Educational Research Bulletin, 1952, 31, 43-47.

Klare, G. R. The measurement of readability. Ames, Ia.: Iowa State University Press, 1963.

Klare, G. R., Sinaiko, H. W., & Stolurow, L. M. The cloze procedure: A convenient readability test for training materials and translations. Arlington, Va.: Institute for Defense Analysis, Science and Technology Division Paper P-660, 1971.

Lee, W. Supra-paragraph prose structure: Its specification, perception, and effects on learning. Psychological Reports, 1965, 17, 135-144.

Lewerenz, A. S. Measurement of the difficulty of reading materials. Educational Research Bulletin, Los Angeles City Schools, 1929, 8, 11-16. Cited by G. R. Klare, The measurement of readability. Ames, Ia.: Iowa State University Press, 1963, P. 40.

Lewerenz, A. S. Vocabulary grade placement of typical newspaper content. Educational Research Bulletin, Los Angeles City Schools, 1930, 10, 4-6. Cited by G. R. Klare. The measurement of readability. Ames, Ia.: Iowa State University Press, 1963.

Lewerenz, A. S. A vocabulary grade placement formula. Journal of Experimental Education, 1935, 3, 236.

Lewerenz, A. S. Selection of reading materials by pupil ability and pupil interest. Elementary English Review, 1939, 16, 151-156.

Lewis, N. Listen please! Clearing House, 1956, 30, 535-536.

Lively, B.A., & Pressey, S.L. A method for measuring the "vocabulary burden" of textbooks. Educational Administration and Supervision, 1923, 9, 389-398.

Long, G. A study on the preparation of an audio-tutorial minicourse. Dissertation Abstracts International, 1970, 31, 1690-1691.

Lorge, I. Predicting reading difficulty of selections for children. Elementary English Review, 1939, 16, 229-237.

Lorge, I. Predicting readability. Teachers College Record, 1944, 45, 404-409.

Lorge, I. The Lorge and Flesch readability formulae: A correction. School and Society, 1948, 67, 141-142.

Lumsdaine, A. Educational technology, programmed learning, and instructional science. Theories of Learning and Instruction. 63rd Year Book, National Society for the Study of Education, Chicago: University of Chicago Press, 1964, 371-401.

Madden, H.L., & Tupes, E.C. Estimating reading ability level from the AQE General Aptitude Index. PRL-TR-66-1, AD-632 182. Lackland AFB, Texas: Personnel Research Laboratory, Aerospace Medical Division, February, 1966.

McClusky, A.Y. A quantitative analysis of the difficulty of reading materials. Journal of Educational Research, 1934, 28, 276-282.

McLaughlin, G.H. SMOG grading: A new readability formula. Journal of Reading, 1969, 12, 639-646.

Morris, E.C., & Halverson, D. Idea analysis technique. Unpublished. On file at Columbia University Library. 1938. Cited by G.R. Klare, The measurement of readability. Ames, Ia.: Iowa State University Press, 1936.

Nelson, J. The effects of visual-auditory modality preference on learning mode preference in first grade children. Dissertation Abstracts International, 1970, 31, 2219-2220.

Nichols, R. Teaching of listening. Chicago Schools Journal, 1949, 30, 273-278.

Ojemann, R. The reading ability of parents and factors associated with the reading difficulty of parent education materials. In G.D. Stoddard (Ed.) Researchers in Parent Education II. Iowa City, Ia.: University of Iowa, 1934. Cited by J.S. Chall, Readability: An appraisal of research and application. Columbus, Ohio: Ohio State University, Bureau of Educational Research, 1958.

Patty, W.W., & Painter, W.I. A technique for measuring the vocabulary burden of textbooks. Journal of Educational Research, 1931, 24, 127-134.

Phillips, M. Learning materials and their implementation. Review of Educational Research, 1966, 36, 373-379.

Powers, R.D., Sumner, W.A., & Kearl, B.E. A recalculation of four adult readability formulas. Journal of Educational Psychology, 1958, 49, 99-105.

Rankin, E.F., & Culhane, J.W. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1969, 13, 193-198.

Senour, R. A study of the effects of student control of audio taped learning experiences (via the control functions incorporated in the instructional device). Dissertation Abstracts International, 1971, 31, 3351-3352.

Severin, W. The effectiveness of relevant pictures in multiple-channel communications. AV Communications Review, 1967, 15, 386-401.

Siegel, A.I., Barcik, J.D., & Macpherson, D.H. Mass training techniques in civil defense. I. Introductory studies of telephonic adjunct training. Wayne, Pa.: Applied Psychological Services, February 1965.

Siegel, A.I., & Fischl, M.A. Mass training techniques in civil defense. II. A further study of telephonic adjunct training. Wayne, Pa.: Applied Psychological Services, July 1965.

Singer, C. Eye movements, reading comprehension, and vocabulary as effected by varying visual and auditory modalities in college students. Dissertation Abstracts International, 1970, 31, 164.

Smith, E.A., & Kincaid, J.P. Derivation and validation of the Automated Readability Index for use with technical materials. Human Factors, 1970, 12, 457-464.

Smith, E.A. & Senter, R.J. Automated Readability Index. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Division, AMRL-TR-66-220, November 1970.

Spache, G. A new readability formula for primary grade reading materials. Elementary School Journal, 1953, 53, 410-413.

Sticht, T.G. Learning by listening in relation to aptitude, reading, and rate-controlled speech. Presidio of Monterey, Calif.: Human Resources Research Organization, HumRRO Division No. 3, Technical Report 69-23, December, 1969.

Stone, C.R. Measures of simplicity and beginning texts in reading. Journal of Educational Research, 1938, 31, 447-450.

Szalay, T.G. Validation of the Coleman readability formulas. Psychological Reports, 1965, 17, 965-966.

Tannenbaum, P.H., & Greenberg, B.S. Mass communication. In P.R. Farnsworth (Ed.), Annual Review of Psychology, 1968, 19, 351-387.

Taylor, W. Cloze procedure: a new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Taylor, W.L. Recent developments in the use of "cloze procedure." Journalism Quarterly, 1956, 33, 42-46, 99.

Taylor, W.L. "Cloze" readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology, 1957, 41, 19-26.

Thorndike, E.L. Improving the ability to read. Teachers College Record, 1934, 36, 1-19, 123-144, 229-241.

Travers, R.M.W. On the transmission of information to human receivers. Audiovisual Communication Review, 1964, 12, 373-385.

Tribe, E.B. A readability formula for the elementary school based upon the Rinsland vocabulary. Unpublished Doctoral dissertation, University of Oklahoma, 1956. Cited by G.R. Klare, The measurement of readability. Ames, Ia.: Iowa State University Press, 1963, P. 69-70.

Valentine, L.D. Jr., & Vitola, B.M. Comparison of self-motivated Air Force enlistees with draft-motivated enlistees. AFHRL-TR-70-26, AD-713 638. Lackland AFB, Texas: Personnel Research Division, Air Force Human Resources Laboratory, July 1970.

Van Mondfrans, A., & Travers, R. Learning of redundant materials presented through two sensory modalities. Perceptual & Motor Skills, 1964, 19, 743-751.

Van Mondfrans, A., & Travers, R. Paired-associate learning within and across sense modalities and involving simultaneous and sequential presentations. American Educational Research Journal, 1965, 2, 89-99.

Vineberg, R., Sticht, T.G., Taylor, E.N., & Caylor, J.S. Effects of aptitude (AFQT), job experience, and literacy on job performance: summary of HumRRO work units UTILITY and REALISTIC. Presidio of Monterey, Calif.: Human Resources Research Organization, HumRRO Division No. 3, August 1970.

Virag, W. The effectiveness of three instructional modes employed to transmit content to students with different aptitude patterns. Dissertation Abstracts International, 1971, 31, 5520.

Vogel, M., & Washburne, C.W. An objective method of determining grade placement of children's reading material. Elementary School Journal, 1928, 28, 373-381. Cited by C.R. Klare, The measurement of readability, Ames, Ia.: Iowa State University Press, 1963, P. 38-39.

Washburne, C., & Morphett, M.V. Grade placement of children's books. Elementary School Journal, 1938, 38, 355-364.

Washburne, C., & Vogel, M. What books fit what children? School and Society, 1926, 23, 22-24. Cited by G.R. Klare, The measurement of readability, Ames, Ia.: Iowa State University Press, 1963. P. 38-39.

Wheeler, L.R., & Smith, E.H. A practical readability formula for the classroom teacher in the primary grades. Elementary English, 1954, 31, 397-399.

Wheeler, L. R., & Wheeler, V. D. Selecting appropriate reading materials. Elementary English, 1948, 25, 478-489.

Yoakam, G. A. A technique for determining the difficulty of reading materials. Unpublished. University of Pittsburgh, 1939. (Revised 1948.) In G. A. Yoakam, Basal Reading Instruction, New York: McGraw-Hill, 1955. Cited by J. S. Chall, Readability: An appraisal of research and application. Columbus, Ohio: Bureau of Educational Research, Ohio State University, 1958. P. 28.

# APPENDIX A

## Glossary of Readability Measures

| Author(s) | Date | Criterion | Variables | Predicted Value | Formula | Type of Material Studied | Range of Material Studied | Method of comparison with criterion |
|---|---|---|---|---|---|---|---|---|
| Lively & Pressey | 1923 | Judged difficulty | Weighted median index no., based on Thorndike word list | Relative difficulty | Weighted median index number | School, general, scientific | Grade 2-college | Inspection |
| Dolch | 1928 | Grade level assigned by publishers | Five indices of vocabulary "load | ——— | Base judgment of difficulty on "several" of vocabulary factors | Reading texts | Primer-grade 4 | Inspection |
| Washburne & Vogel (Winnetka) | 1928 | Mean tested reading level of those liking book | $X_2$: Number of different words in 1000 words<br>$X_3$: Number of prepositions in 1000 words<br>$X_4$: Number of words not on Thorndike list of 10,000 words<br>$X_5$: Number of simple sentences in 75 sentences | Reading score on Stanford Achievement Test ($X_1$) | $X_1 = .085X_2 + .101X_3 + .604X_4 - .411X_5 + 17.43$ | Children's library books | Grades 3-9 | Correlation .845 (This value is confounded by effects of popularity of books) |
| Lewerenz | 1929 | Order of presentation on Stanford Achievement Test | Percentage of words beginning with W, H, B (easy) I, and E (hard) | Grade level | Mean of tabled values for each of variables | Stanford Achievement Test paragraphs | Grade 2-college | Graphic |
| Lewerenz | 1930<br>1935<br>1939 | ——— | 1) Ratio of Anglo-Saxon to Greek and Roman words (difficulty)<br>2) Ratio of words in "Clark's first 500" to total different words (diversity)<br>3) Estimate of image bearing or sensory words (interest)<br>4) Polysyllabic word count<br>5) Vocabulary mass (number of different words in sample that are not on Clark's list of 500 common words | Grade level | Mean of tabled values for each of variables | ——— | ——— | ——— |
| Johnson | 1930 | Grade level assigned by publishers | Percentage of polysyllable words | Grade level | Tabled value of percentage of polysyllables | Wide variety | Primer to grade 8 | Inspection |
| Patty & Painter | 1931 | None (relied on validity of Thorndike word list) | 1) Thorndike word list index numbers<br>2) Number of different words in sample | Relative difficulty | Index Number = $\frac{A.W.W.V.}{12}$<br>A.W.W.V. = mean Thorndike Index Number<br>R = number of different words in sample | School texts | Grades 9-12 | ——— |

| Author(s) | Date | Criterion | Variables | Predicted Value | Formula | Type of Material Studied | Range of Material Studied | Method of comparison with criterion |
|---|---|---|---|---|---|---|---|---|
| Ojemann | 1934 | $X_{C50}$ | 6 measures of vocabulary difficulty, 8 measures of structure, 4 qualitative factors | ——— | 16 passages are presented in order of difficulty, to be compared with material to be tested | Adult education | Grade 6-college | Quantitative factors: correlation. Individual factors up to .60 Qualitative factors: inspection. |
| Dale & Tyler | 1934 | Comprehension score of adults of low reading level | $X_2$: Number of different technical terms<br>$X_3$: Number of different hard non-technical words<br>$X_4$: Number of indeterminate clauses | Percent of adults of low RGL (3-5) who will understand passage ($X_1$) | $X_1 = -9.4X_2 - .4X_3 + 2.2X_4 + 114.4 \pm 9.0$ | Adult health education material | Below grade 8 | Correlation .511 |
| McClusky | 1934 | Reading speed, comprehension score | Vocabulary, sentence structure | ——— | No formula. Description of vocabulary and sentence structure characteristics in easy and hard materials. | Wide variety | Above grade 8 | Inspection |
| Thorndike | 1934 | Vocabulary difficulty | Number of words not on Thorndike list in 10,000 word sample | Grade level | Tabled value of number of words not on Thorndike list. | General fiction and non-fiction | Grade 4-9 | ——— |
| Gray and Leary | 1935 | Mean comprehension test score of adults | $X_2$: Number of different words not on Dale List of 769<br>$X_5$: Number of personal pronouns<br>$X_6$: Mean sentence length in words<br>$X_7$: Percent different words<br>$X_8$: Number of prepositional phrases | Mean comprehension test score of adults ($X_1$) | $X_1 = .01029X_2 + .009012X_5 - .02094X_6 - .033.3X_7 - .01485X_8 + 3.774$ | General adult | Grade 2-college | Correlation .6435 |
| Washburne & Morphett (revision of Vogel & Washburne, 1928) | 1938 | Mean tested reading level of those liking book. Also teacher judgment at grades 1-2. | $X_2$: Number of different words per 1000 words<br>$X_3$: Number of different words per 1000 not in Thorndike's most common 1500 words<br>$X_4$: Number of simple sentences in 75 sentences | Grade level ($X_1$) | $X_1 = .00255X_2 + .0458X_3 - .0307X_4 + 1.294$ | Children's library books | Grades 1-9 | Correlation .86 (This value is confounded by effects of popularity of books) |

| Author(s) | Date | Criterion | Variables | Predicted Value | Formula | Type of Material Studied | Range of Material Studied | Method of comparison with criterion |
|---|---|---|---|---|---|---|---|---|
| Stone | 1938 | _____ | Ratio of new words to total words, percent sentences complete on one line | Relative difficulty | Relative difficulty based on variables stated. | Reading texts | Grade 1 | _____ |
| DeLong | 1938 | _____ | Percentage of words in various levels of Stone's word list. | Six levels of relative difficulty | Tabled value of percentages of words at various levels of Stone's word list | Reading texts | Pre-primer to grade 2 | _____ |
| Lorge | 1939<br>1944<br>1948 | McCall-Crabbs test norms | $X_2$: Percent of words outside Dale list of 769 words<br>$X_6$: Mean sentence length<br>$X_8$: Percent prepositional phrases | Reading grade level needed to answer 50% of questions correctly ($C_{50}$) | $C_{50} = .10X_2 + .06X_6 + .10X_8 + 1.99$ | McCall-Crabbs test passages | Grades 3-12 | Correlation .67 |
| Morris & Halverson | 1938 | _____ | 4 Qualitative vocabulary classes:<br>I) words learned early in life<br>II) localisms<br>III) concrete words<br>IV) abstract words | Relative difficulty | Tabled norms relate count of words in 4 classes to difficulty | _____ | _____ | Multiple correlation of .74 was found between classes I, III, IV, and McCall-Crabbs test norms (Lorge, 1939) |
| Yoakam | 1939 | _____ | Vocabulary index: based on Thorndike word list. Words between 4th and 20th thousand, index is the number of thousand in which word appears. Above 20th thousand index = 20. | Grade level | Mean index number comparison with tabulated values | School texts | 2nd -14th grade | Inspection |
| Kessler | 1941 | Judged difficulty | Mean sentence length in words<br>Mean number of different hard words per 100 words | Grade level | Factors compared to Gray and Leary standards independently | 35 textbooks | Grade 2 - college | Inspection |
| Flesch | 1943 | Judgment and McCall-Crabbs test norms | $X_s$: Mean sentence length<br>$X_m$: Number of affixes per 100 words<br>$X_h$: Number of personal references per 100 words | Reading grade level needed to answer 75% of questions correctly ($C_{75}$) (correction for high grade levels) | $C_{75} = .1338X_s + .0645X_m - .0659X_h + 4.2498$<br>Later corrected to;<br>$C_{75} = .07X_m + .07X_s - .05X_h + 3.27$ | Adult magazines and McCall-Crabbs passages | Grades 3-12 | Correlation .74 |

| Author(s) | Date | Criterion | Variables | Predicted Value | Formula | Type of Material Studied | Range of Material Studied | Method of comparison with criterion |
|---|---|---|---|---|---|---|---|---|
| Flesch (revision) | 1948 | Judgment and McCall-Crabbs test norms | wl = Mean word length (syllables per 100 words)<br>sl = Mean sentence length<br>pw = Number of personal words per 100 words<br>ps = Number of personal sentences per 100 sentences | Relative difficulty, relative interest | Reading Ease = $206.835 - .846_{wl} - 1.015_{sl}$<br>Human Interest = $3.635_{pw} + .314_{ps}$ | McCall-Crabbs passages | Grades 3-12 | Correlation<br>R.E.: .70<br>H.I.: .43 |
| Dale & Chall | 1948 | McCall-Crabbs test norms, comprehension test scores | $X_1$: Percent of words not on Dale list of 3000 common words<br>$X_2$: Mean sentence length in words | Reading grade level needed to answer 50% of questions correctly ($C_{50}$) | $C_{50} = .1579X_1 + .0496X_2 + 3.6365$ | McCall-Crabbs passages, school material, health articles | Grades 3-12, 3-16 for health passages | Correlation .70 |
| Dolch | 1948 | Grade level assigned by publishers | $X_1$: Median sentence length<br>$X_2$: 90th percentile sentence length<br>$X_3$: Percent hard words (not on Dolch list of 1000 words) | Grade level | Look up three variables in tables, average the tabled values. | Reading texts | Grades 1-6 | Inspection |
| Wheeler & Wheeler | 1948 | None (relied on validity of Thorndike word list) | Percent of words in each thousand of Thorndike 20,000 word list | Instructional, Independent, and average grade level | Tabulate words in passages by Thorndike thousands. For instructional level, 90% of words must be below student's RGL. For independent level 95%. | ——— | ——— | ——— |
| Flesch | 1950 | Comprehension test scores | wl: Word length (syllables per 100 words)<br>dw: percentage of "definite words." | R: Level of abstraction" on arbitrary scale | $R = 168.095 + .532_{dw} - .811_{wl}$ | School texts | Grades 3-12 | Correlation .72 |
| Farr, Jenkins, & Patterson | 1951 | Flesch reading ease score | nosw: Number of one-syllable words<br>sl: Mean sentence length | Relative difficulty | New Reading Ease Index = $1.599_{nosw} - 1.015_{sl} - 31.517$ | Adult, public relations | Full range of Flesch "reading ease" values | Correlation .95 |
| Gunning | 1952 | Flesch scores, McCall-Crabbs norms, and judgment | Mean sentence length, percent of words of 3 or more syllables | Fog Index (the reading grade level needed to read and understand material) | Fog Index = .4 x (mean sentence length + percent of words of 3 or more syllables) | General adult and school | Grades 6-12 | Inspection |

| Author(s) | Date | Criterion | Variables | Predicted Value | Formula | Type of Material Studied | Range of Material Studied | Method of comparison with criterion |
|---|---|---|---|---|---|---|---|---|
| McElroy (AF Manual 11-3) | 1953 | _____ | Easy words (pronounced in 1-2 words), assigned value 1 hard words (pronounced in 3 or more syllables), assigned value of 3 | Fog Count (arbitrary value) and reading grade level | Fog Count per sentence = sum of 1's and 3's in sentence RGL = sum of 1's and 3's divided by number of sentences. If value is over 20, divide by 2. If below 20, subtract 2, then divide by 2. | _____ | _____ | _____ |
| Forbes & Cottle | 1953 | Mean value of five existing readability formulas | Mean Thorndike vocabulary difficulty index, words in 4th thousand and above. | Reading grade level | Look up mean vocabulary index in tables to obtain RGL | Standardized tests | Grade 5-college | Correlation .96 |
| Spache | 1953 | Grade level assigned by publishers | $X_1$: Mean sentence length<br>$X_2$: Percentage of words not on Dale list of 769 easy words | Grade level | Grade level = $.141X_1 + .086X_2 + .839$ | Grade school tests | Grades 1-3 | Correlation .82 |
| Flesch | 1954 | Observed characteristics of easy and hard material | - Number of references to human beings, etc., per 100 words<br>- Number of indications of voice communication per 100 words. | Arbitrary scales of realism ("r") and energy ("e") | "r" score: tabled value corresponding to observed number of references to human beings, etc.<br>"e" score: tabled value corresponding to observed number of indications of voice communication | General adult material | Adult | Inspection |
| Wheeler & Smith | 1954 | Grade level assigned by publishers | Mean length of "units" (similar to sentences), percentage of multi-syllable words. | Reading grade level | Index = 10 (mean length of units x percent multi-syllable words). Look up index in table to obtain RGL | School texts | Grades 1-4 | Inspection |
| Tribe | 1956 | McCall-Crabbs test scores (revised form) | $X_1$: Mean sentence length<br>$X_5$: Percentage of words not on Rinsland word list, multiplied by 100. | Reading grade level needed to answer 50% of questions correctly ($C_{50}$) | $C_{50} = .0719X_1 + .1043X_5 + 2.9347$<br>-Correction is then applied through reference to table of Dale and Chall (1948) | McCall-Crabbs passages, 1950 revision | Grades 3-12 | _____ |
| Gillie | 1957 | Flesch "level of abstraction" score. | $X_1$: Number of finite verbs per 200 words<br>$X_2$: Number of definite articles with nouns per 200 words<br>$X_3$: Number of nouns of abstraction per 200 words. | Abstraction level (arbitrary scale) | Abstraction level = $36 + X_2 + X_1 - (2X_3)$ | School texts, general adult material | Grade 4-college | Correlation .83 |

| Author(s) | Date | Criterion | Variables | Predicted Value | Formula | Type of Material Studied | Range of Material Studied | Method of comparison with criterion |
|---|---|---|---|---|---|---|---|---|
| Stone | 1957 | Grade level assigned by publishers | $X_1$: Number of finite verbs per 200 words<br>$X_2$: Number of definite articles with nouns per 200 words<br>$X_3$: Number of nouns of abstraction per 200 words | Abstraction level (arbitrary scale) | Abstraction level = $36 \cdot X_2 - X_1 - (2X_3)$ | School texts, general adult material | Grade 4-college | Correlation .83 |
| Powers, Sumner, & Kearl | 1958 | McCall-Crabbs Test Lessons, 1950 revision | Recalculation of four existing formulas: Flesch Reading Ease, Dale-Chall, Farr, Jenkins, and Paterson, and Gunning. Variables and unchanged. Coefficients altered as appropriate | $X_{C50}$ | Flesch: $X_{C50} = -2.2029 + .0778$ (sentence length) + .0445 (syllables per 100 words) | McCall-Crabbs passages, 1950 revision | Grades 3-12 | Correlation .64 |
| | | | | $X_{C50}$ | Dale-Chall: $X_{c50} = 3.2672 + .0596$ (sentence length) - .0648 (percent of words not on Dale list) | McCall-Crabbs passages, 1950 revision | Grades 3-12 | Correlation .71 |
| | | | | $X_{C50}$ | Farr-Jenkins-Paterson: $X_{C50} = 8.4355 + .0923$ (sentence length) - .0648 (percent of one syllable words) | McCall-Crabbs passages, 1950 revision | Grades 3-12 | Correlation .58 |
| | | | | $X_{C50}$ | Gunning: $X_{c50} = 3.0680 + .0877$ (sentence length) + .0984 (percent of one syllable words) | McCall-Crabbs passages, 1950 revision | Grades 3-12 | Correlation .59 |
| Flesch | 1958 | Observed characteristics of written material | Count of evidences of informality: capitalization, underlining, punctuation, symbols, beginnings or ends of paragraphs, etc. | Arbitrary scale of "formality-popularity" | Find total number of evidences of informality per sample, use table to determine formality-popularity category. | Popular to scholarly periodicals | Adult | Inspection |
| Bloomer | 1959 | Grade level assigned by publishers | - Number of words per modifier<br>- Sound complexity of modifiers | Level of abstraction | Grade level predicted from the variables listed. Procedure not described. | Reading texts | Grades 1-6 | Correlation .78 |
| Jacobson | 1965 | Mean number of words stated as not-understood by readers | $X_1$: Mean number of words per independent clause<br>$X_2$: Concentration of mathematical terms<br>$X_3$: Concentration of words outside of most common 6,000 on Thorndike list<br>$X_4$: Concentration of words not on Power's list of essential scientific terms. | Y: Mean number of words not known per sample | Physics texts: $Y = -.0013 + 29.7059X_2 + 21.119X_3 + 35.0029X_4$<br>Chemistry texts: $Y = .003 + .1706X_1 + 13.7231X_2 - 43.7262X_3 - 2.3577X_4$ | Physics & Chemistry texts | High-school & college | Correlation<br>Physics texts: .70<br>Chemistry texts: .67 |